LACIE-C00202
JSC-11669

LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)

NASA NOAA USDA

Level 3 Baseline

IMAGE-100/HYBRID SYSTEM PROCEDURE 1 REQUIREMENTS

NOTICE: THIS IS A BASELINED LEVEL 3 DOCUMENT CONTROLLED BY THE LACIE LEVEL 3 CHANGE CONTROL BOARD. ANY PROPOSED CHANGES SHOULD BE DOCUMENTED ON AN RECP FORM AND TRANSMITTED TO R. B. MACDONALD, LACIE MANAGER, NASA-JSC, CODE SF, HOUSTON, TEXAS 77058.

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

JANUARY 1977

REVISIONS

REV LTR	CHANGE NO.	DESCRIPTION	DATE
		BASELINE ISSUE (Reference CCBD 7M0001 dated January 6, 1977)	1/6/77

LIST OF EFFECTIVE PAGES

The current status of all pages in this document is as shown below:

Page No.	LACIE Change Date	Authorizing CCBD No.
ii through v	Baseline 1-6-77	7M0001
vi	11-17-77	7M0028
1 and 2	Baseline 1-6-77	7M0001
3 through 4a	11-17-77	7M0027
5 through 10	Baseline 1-6-77	7M0001
11 through 15	5-24-77 and 11-17-77	7M0010 7M0028

LARGE AREA CROP INVENTORY EXPERIMENT
(LACIE)

IMAGE-100/HYBRID SYSTEM
PROCEDURE 1 REQUIREMENTS

APPROVED BY:

for R. B. MacDonald
LACIE Manager

JANUARY 1977

ACRONYMS AND ABBREVIATIONS

CAMS	Classification and Mensuration Subsystem
CAS	Crop Assessment Subsystem
CIR	Color infrared
COM	Computer output microfiche
CRT	Cathode ray tube
DO	Designated other
DTERM	Data terminal
DU	Designated unidentifiable
ERIPS	Earth Resources Interactive Processing System
ID	Identification
IMAGE-100	Interactive Multispectral Image Analysis System, Model 100
LACIE	Large Area Crop Inventory Experiment
Landsat	Land Satellite
Pixels	Picture elements

CONTENTS

1. INTRODUCTION

Procedure 1 is an improved classification scheme which should solve problems encountered in Large Area Crop Inventory Experiment (LACIE) Phases 1 and 2. The analysis uses as a starting point a preselected set of picture elements (pixels) or dots which are defined as starting vectors or are labeled as training data. The segment is clustered using a randomized set of the dots as starting vectors. The resulting clusters are labeled as wheat or nonwheat automatically using the labeled dots. The labeled subclass cluster statistics are used to classify the total segment. The classified dots or pixels are checked for labeling error and a classifier bias correction is computed and applied; this set of labeled dots is retained and used on the next acquisition prior to interaction analysis by the analyst.

The following requirements are defined for the Interactive Multispectral Image Analysis System (IMAGE-100)/IBM 360/75 Hybrid System.

2. HYBRID SYSTEM

2.1 GENERAL DISCUSSION

In a normal processing flow for the hybrid system, when the analyst receives the first imagery for the segment, it is studied and DO/DU fields delineated for the segment. These fields may be entered into the data base via Dell Foster card output or interactively.

The user then interactively labels the grid intersections, up to 209, aided by spectral displays which allow him to simultaneously locate and view the dot in the spatial and spectral domains. This labeling may occur offline and the data base updated via cards. Labeling occurs in two sequences, once for Type 1 (labeling and

starting vectors) and once for Type 2 (bias correction vectors). These sequences may occur concurrently or separately (labeling via cursor or console keyboard). After labeling and typing are completed, the information is stored in the data base, where it is off-loaded daily. The off-loading would consist of dumping those files which had been updated that day, selectively reformatting, and punching the information onto cards in the proper format to update the LACIE (IBM 360/75) Fields and Dot Data Base. These cards also contain the information identifying the acquisition to be used in processing.

After the 360/75 data base has been initialized, a LACIE batch run is executed which retrieves the Dot Data Base information and executes the functions of clustering, automatic cluster labeling, classification, class summary generation (including bias correction if bias correction dots were available), and acquisition separability computation.

Outputs from this processing to the IMAGE-100 are conditional cluster map, unconditional cluster map, classification map, and classification summary. The maps are on a single tape [data terminal (DTERM)] and the reports are on the computer output microfiche (COM) tape. The summary information can be retrieved from the COM tape or regenerated based on the Dot Data Base, designated other/designated unidentifiable (DO/DU) field def., and the class map.

When these products are available, the analyst interactively evaluates the results of the LACIE batch run. At this time, the user can access any of the labeling modules or view any of the above-mentioned products. If only Type 1 dots were labeled initially, he will want to label bias correction dots (Type 2) and compute the classification summary information. Other functions which might be accessed during this session would be to

a. Display the conditional clusters (up to 60) and alarm all dots or subsets of dots (WII, T2, N, etc.) which fall within a particular cluster

b. Relabel current dots or select additional dots and label

c. Relabel cluster and create new class map by relabeling all pixels assigned to this cluster with the new cluster label

d. Display spectral/trajectory plots of dots and/or subsets of dots

e. Relabel Type 2 dots and recompute summary information

f. Manually enter summary and map information

g. Update DO/DU field definitions

After all verification and relabeling have been completed, the information is stored in the IMAGE-100 Data Base and subsequently transmitted to the LACIE Data Base. The user may perform interactive rework on the 360/75. Output of this mode would be hardcopy summary and cluster reports and a DTERM tape containing cluster and classification maps. These are then returned to the IMAGE-100 for the verification process.

After the final data base update has occurred for this acquisition, the system is in a "wait state" until the arrival of the next acquisition for this segment. When the new acquisition arrives, the batch processing is initiated on the 360/75 and products for the IMAGE-100 generated.

The user then has the same options for labeling and relabeling as in the single acquisition case. Additionally, when the number of acquisitions used for processing reaches four, the user is required to select three of the four for use in the next batch processing along with the latest acquisition. This selection is based on separability information generated during batch processing.

2.2 DATA BASE

The Hybrid System shall include a data base with the capacity to store 4-band MSS Landsat data for a maximum of 6 acquisitions per segment for a maximum of approximately 200 segments. The entire data base need not be available on-line; it will be satisfactory if it is divided into blocks each of which contains all of the information for about 100 segments such that one block is on-line at a time.

The data base shall contain the class summary information and the cluster and class maps for the most recent ERIPS job that resulted in an acceptable category proportion estimate for each segment. In addition, the data base shall provide storage for the same information from a more recent ERIPS job for each segment that is awaiting evaluation.

The data base shall provide storage for a maximum of 50 fields per segment where a field is a polygon with not more than 10 vertices. The data base shall also provide storage for the following set of dot related data for each segment: analyst label and type, Landsat data in both MSS and Tasselled Cap coordinates for each acquisition, random sequence and grid numbers, and the most recent classifier label. For purposes of unloading analyst dot labels and types and field definitions for a segment, the I-100 data base shall contain an analyst input set 1/set 2 flag for each segment to indicate which of the two sets of dots and fields in the LACIE data base is to be updated with the result of this unload.

The image display gain and bias that is computed at GSFC and stored in the image header shall be stored in the data base for each acquisition.

In addition to a capability to enter dot types and labels interactively at the analyst console and subsequently store this information permanently in the data base, there shall be a batch capability to enter and update dot types and labels on a segment-by-segment basis via card input.

It should be assumed that the execution of batch jobs requiring access to the data base will be scheduled during a time period when no interactive analysis is being performed; thus, no conflicts will occur between batch and interactive tasks that require access to the same data base files.

2.3 GENERAL DISPLAY CAPABILITIES

The display capabilities are closely linked with the labeling capabilities. This section will attempt to address those "universal" capabilities without the special links to the random dot sequence and the labeling functions. These will be addressed in a section on labeling.

The system shall be able to display one acquisition in the color infrared (CIR) image type display in conjunction with computer-generated maps and/or spectral and trajectory plots. If spectral plots are displayed, one must correspond to the acquisition being displayed in CIR image form. The user shall be able to select the type of display and determine its size (within limits), position on the screen and acquisition for which each display is generated. Classification and cluster maps may be displayed independently or overlaid on the CIR such that they can be "seen through." A grid locating the 209 intersections shall be available for overlaying on any image.

Spectral plots of the 209 dots for up to four acquisitions and a trajectory plot shall be able to be viewed simultaneously and in conjunction with the CIR image. Details of individual spectral and trajectory plots will be included in another section.

A cross-alarming capability shall exist between spectral plots and the CIR. If a cluster or class map is overlaid on the CIR, the alarm should still be visible. This alarm is triggered by cursor on either the spectral plot or the CIR image or by designation of a dot number. The capability to turn the alarm off and on for viewing the spectral value of the dot shall exist.

If a dot or element is cursored on the spectral plot, all dots having that same vector in that acquisition will be alarmed in the CIR image. However, if a dot is cursored on the CIR image, the user will have the option to alarm in both the CIR image and spectral plot(s) or either one exclusively. If more than one spectral plot is displayed, the correct location shall be alarmed in all of those displayed.

A report of dot grid number, random sequence number, analyst label, green number, type, and classification, if available, shall be optionally generated when a dot is cursored on the CIR image.

The user shall also have the capability to enter a dot grid number and have that location alarmed in the CIR image and its corresponding position in the spectral plot. This capability should also be available for multiple dot numbers or dot groups (such as wheat, nonwheat, etc.).

The same three modes should be available for causing a trajectory plot to be generated. If a dot number is entered, the trajectory plot utilizing all acquisitions on the data base or the subset of

these acquisitions specified by the user is generated. The same output is created if the dot is cursored on the CIR image. If a "dot" is cursored on the spectral plot, all dots having that vector for that acquisition are alarmed in the CIR image. Then a trajectory plot for each one is generated. Dots may be addressed for spectral display by type, label, classification label, green number or any logical combination of the above.

All displays, where applicable, can be shown in any "channel" subset selected by the user or from either the Land Satellite (Landsat) or the Kauth coordinate system. The option to normalize the display axes in accordance with the "soil line" shall be provided.

2.4 DOT LABELING

The user will have the capability to label dots by a number of methods. The user may request the first m (m = 1, 209) dots in the random sequence not lying in DO or DU areas to be displayed (shown on the grid, alarmed on the screen). The analyst may then communicate labels for as many of these as he wishes. Labels may be communicated by cursoring dots and assigning a single label to all. Labeled dots shall be distinguishable from unlabeled dots. Each category shall also be distinctly identifiable. The user may then cursor more dots and assign another label. Dots may be cursored for labeling on either the CIR image or the spectral plot. If an area is cursored on the spectral plot, all dots (grid intersections) whose vector lies in that area and are alarmed for labeling receive that label. Labels may also be typed in by dot number. The cross-alarming capabilities described in section 2.3 will enable a user to interact between the spatial and spectral displays. Basically, the user can cross-alarm and accumulate dots until a decision has been made, then a label can be applied to the collection of dots accumulated or to single dots.

After the set of dots being viewed has been assigned labels, the user can either request additional dots from the random sequence or he may designate the type for all dots in the set. If all dots were labeled together, a "division point"[1] between starting vectors and labeling vectors and bias correction vectors can be entered.

After the Type 1 dots have been labeled, the user can either request the next p dots from the random sequence or the first p unlabeled dots from the random sequence to be displayed for labeling. This second set of dots will generally be the Type 2 dots.

When a user is labeling Type 2 dots, the option shall exist for the dots to be selected in accordance with the classified proportions for the current processing. That is, the user requests 100 Type 2 dots. The class summary information has been loaded. If, for instance, the segment was classified 30 percent wheat and 70 percent nonwheat, the first 30 Type 2 dots in the random sequence which were classified wheat and the first 70 Type 2 dots which were classified nonwheat are selected. When these are alarmed for labeling, the user is not shown which are the classified wheat and which are classified nonwheat. This option is inoperative if current classification information is unavailable.

The same capabilities described to aid in the decision process for Type 1 dots should be available for Type 2 dots. The labeling of Type 1 and Type 2 dots may occur in the same interactive session or in separate sessions. The dot labels will be transmitted to the LACIE Data Base in the following order: all Type 1 dots in the random sequence order, then all Type 2 in the

---

[1] The first m labeled dots are starting vectors; the first n labeled dots are labeling vectors, where $m \leq n < 209$; the remaining labeled dots are bias correction vectors.

random sequence. Any time the user changes a label or a type, an update to both the IMAGE-100 and the LACIE data bases will occur. Dot labels may be changed at any time during an IMAGE-100 terminal session. If the user adds bias correction vectors (Type 2 dots) or relabels bias correction vectors, he can then require a new bias correction and dot summary to be generated. The details of this computation and summary are included in the "Class Summary Changes" RECP, ASVB-098.

After each complete session in labeling (the user stores information on data base), a final report shall be generated for that segment listing each dot grid number, random sequence number, analyst label, type (i.e., starting, labeling or bias correction) and last classification.

2.5 INTERACTIVE TRAJECTORY AND SCATTER PLOTS

Analyst decides where the plot for any of six acquisitions is positioned on the screen. There will be default positions for LACIE.

2.6 TRAJECTORY AND SCATTER PLOTS

User specifies a dot by number or cursor for plotting. Each acquisition is plotted on a different (user specified) theme track. (A line will optionally be drawn from acquisition n to acquisition m on the theme for acquisition n.)

Each plot will be generated on a Greenness/Brightness axis or two (user-specified) channels.

The scatter plot is simply an accumulation of trajectory plots on one theme for groups of dots and/or cluster means (i.e., classified W, analyst labeled W, etc., constituting display options available to the analyst). The acquisition is

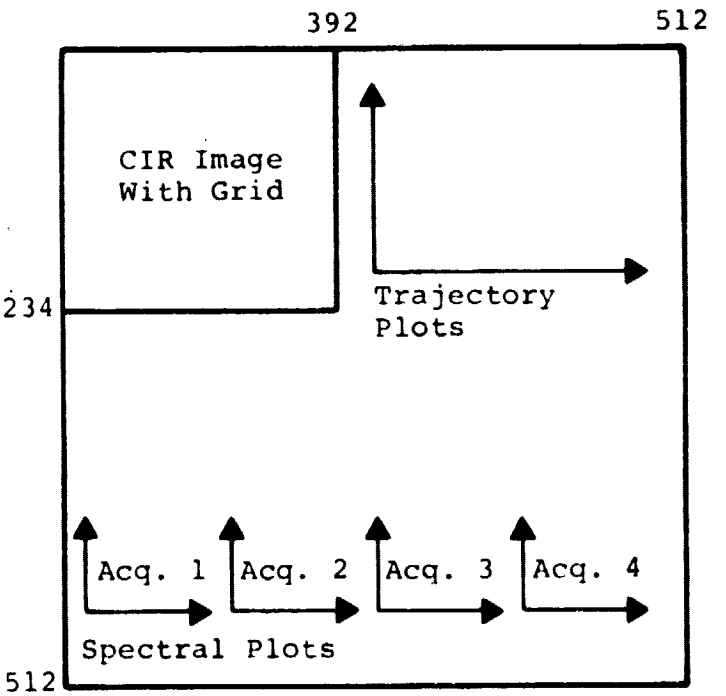
analyst specified. The same acquisition may be displayed at different locations on the console. Logical operations are required. The green number will be printed for any analyst specified dot, as will the current analyst label and machine classification if any has been assigned.

The grid defining the dots will be optionally placed on a theme over the CIR image.

The module will allow the analyst to change the label of a specified dot in a temporary file and reexercise the automatic bias correction logic. The temporary file may be saved in another module as permanent.

A different module is executed for displaying the image on the screen.

CRT Display



2.7 CLUSTER LABELING

The user shall have the capability to relabel a cluster. This implies the user will possess several pieces of information.

a. **How the cluster was originally labeled**

b. **The exact location of the cluster (i.e., which pixels are in it)**

c. **How the dots in that cluster were classified**

d. **How the dots in that cluster were labeled**

e. **How the dots in that cluster were used (i.e., were they labeling dots; if so, were they starting dots; or were they bias correction dots)**

Item a: Can be retrieved from the LACIE Cluster Labeling Menu which will appear on the COM tape. The analyst may have the information by means of a hardcopy report generated from that tape or from IMAGE-100 processing of the COM tape.

Item b: The cluster map will give the exact location of clusters. The capability to produce a dot summary of all dots or alarm all dots lying within user-specified clusters shall be provided.

Item c: An overlay of all wheat and nonwheat classified dots would accomplish this.

Item d: An overlay of all dots labeled (1) wheat and (2) nonwheat.

Item e: An overlay of Type 1 and Type 2 dots would accomplish this. Logical operations between these overlays would enable detailed viewing of dots in clusters. Spectral plots of these subsets of the dots and cluster means may be requested.

After all this information has been assembled and displayed to the analyst's specifications, the analyst can relabel the cluster if necessary.

If the cluster is relabeled, the area on the class map represented by that cluster should be replaced with the new symbol. This also entails re-creating the class summary information and altering and classification for the dots falling in that cluster to correspond to the new label. The bias correction can then be recomputed after all cluster relabeling has been done, or upon command after each relabeling.

Also, using this information the analyst can determine the correct label and interactively rework the segment on the ERIPS and supply results to the IMAGE-100. It should be noted that when an interactive 360/75 rework occurs, no COM tape is generated.

2.8 BIAS CORRECTION AND SUMMARY

Because the user is allowed to label bias correction vectors, alter current bias correction vector labels, and change the labels of clusters, the capability to recompute the proportion of each category, apply a bias correction and regenerate a summary is required.

The labels from the dot data file, the classification map and current classification summary information are necessary for performing these computations.

2.9 INTERFACE CAPABILITIES

For this Hybrid System to function, certain information must be routinely passed between the Image 100 and the LACIE system.

First, sample segment imagery in digital form must be unloaded from the LACIE Data Base and transferred to the Image 100 Data Base. This data will be in the form of a 4-channel/acquisition image in Universal format on a LACIE image unload tape. The

header record of each image contains the green value of the soil line for that acquisition as computed at the time that it was loaded into the LACIE Data Base. For details on the format of the LACIE image unload tape see Ref. 1.

Second, the Image 100 must output DO/DU field definitions and dot types and labels for input to the LACIE system via cards. Field definitions should be in the current LACIE Field Data Base format as described in Ref. 2. Dot labels and types should be in the current LACIE Dot Data Base format as described in Ref. 3. Whenever a dot or field unload is performed, it should be a complete unload for the segment in question. Since the LACIE Data Base will be able to hold two sets of dots and fields, the Image 100 user should be able to input a set 1/set 2 indicator for his dot/field unload (see Refs. 4 and 5). The previous value of this set indicator should be stored in the Image 100 Data Base and displayed to the user just prior to initiating the next dot/field unload. The dot/field unloads should be accumulated on disk as they are initiated and then output to cards via a batch job upon request. Such a request might typically be made separately by each analyst who works a segment.

Third, the LACIE system will output cluster maps and classification maps as single channel universal format images on the DTRM tape. The Image 100 will be required to read, store in its data base, display and manipulate these maps. Standard color tables specifying correspondences between color and cluster number and between color and category label will be used to code each map. In reading these maps it will be necessary to skip over the color key information that appears as part of each image on a DTRM tape (for details see Refs. 6 and 7).

Fourth, the LACIE system will output to the CAMS/CAS interface tape other results from the processing of a segment such as: dates of acquisitions used in batch job, classification summary information; cluster number, labels and statistics; category separability; and sun angle normalization flag. These items must be read from the CAMS/CAS tape (Refs. 8 and 9) and stored in the Image 100 Data Base. The requirements for using this information have been discussed in previous sections of this document.

2.10 REWORK CAPABILITY

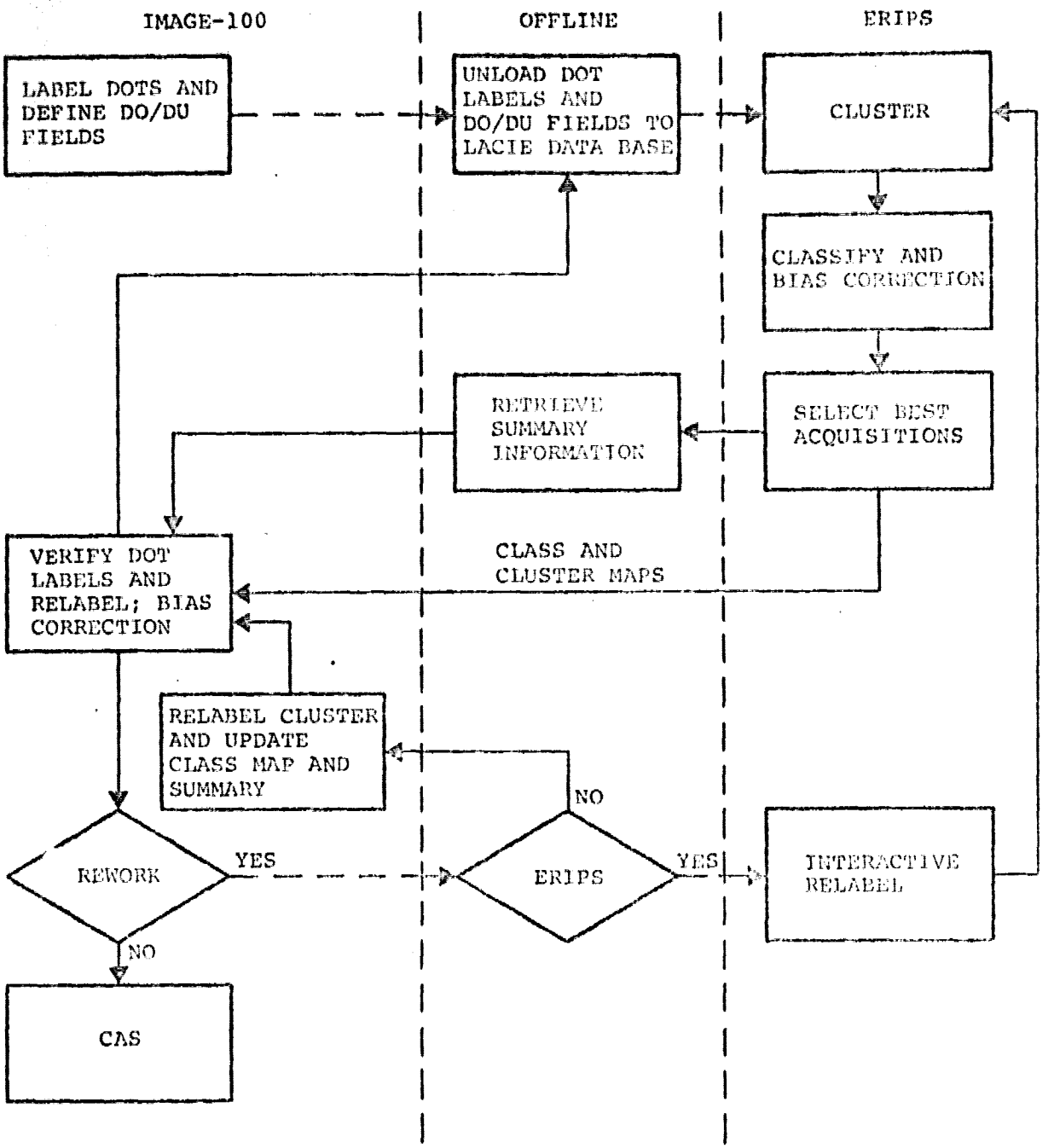
Cluster statistics are required to assist in the relabeling function and in the detection of mixed clusters. These statistics may be generated from a cluster map or by reading the statistics from either the CAMS/CAS Interface Tape or the COM Tape. The cluster means will also be used in generation of spectral plots.

The capability to compute the green number for each cluster mean for any analyst-specified acquisitions shall be provided.

With these statistics, distance tables may be constructed between cluster and labeling vectors and conditional cluster maps generated using various thresholds and distances for the k nearest neighbors. The user shall have the capability to display these conditional cluster maps.

Additionally, these tables may be used for the detection of mixed clusters. These map generation capabilities in conjunction with the cluster relabeling capabilities (i.e., merging maps and reworking classification summary) allow the user to create a product reflective of the estimate without requiring additional 360/75 processing time.

HYBRID SYSTEM



References:

1. LACIE 6 Detail Design Specification, Appendix D: Image Unload Tape Format, prepared by Dept. HR4/IBM, c. May 1977.

2. Large Area Crop Inventory Experiment (LACIE) User's Guide, Vol. 1, Sec. 11, Fields Data Base Transaction System, Insertion 9, May 25, 1977.

3. LACIE 6 Detail Design Specification, Sec. 3.9 - Dot Data Base Utility, prepared by Dept. HR4/IBM, c. May 1977.

4. RECP 7M0021, Field and Dot Unload, July 21, 1977.

5. LACIE 7 Design Review (Part B), Two Sets of Fields and Dots, Aug. 5, 1977, copy of IBM presentation.

6. RECP 7M0023, Color Key for Cluster Map, July 21, 1977.

7. LACIE 7 Design Review (Part A), Color Key for Cluster Maps, Aug. 29, 1977, copy of IBM presentation.

8. LACIE 7 Design Review, (Part A), Aug. 29, 1977, copy of presentation by IBM.

9. CAMS/CAS Utility Interface Data Sets, draft by Barbara Duprey, IBM, Oct. 6, 1977.